분산분석에서 사후분석방법

2018년 1학기

변 종석

1. 개념

분산분석후 모수인자들의 수준간 유의한 평균 차이가 존재한다고 인정되는 경우 구체적으로 어느 수준에서 모평균의 차이가 발생하는 지를 분석하는 방법

① 각 수준에서의 개별 모평균 추정

② 여러 수준간 모평균 차이의 동시 검정

i) 개별 평균간 차이 검정

ii) Inference about structured Means

2. 개별 평균간 차이에 대한 검정

1) 최소유의차 검정 (Least Significant Difference : LSD)

① 개념 : 수준 수가 k개인 경우 두 수준간 비교 가능한 모든 $k(k-1)/2$ 쌍에 대하여 전통적인 t-검정을 통하여 두 수준간 평균 차이 비교하는 방법으로 하나의 기각값만을 이용하여 비교

② 표현식

· 일반적인 t-검정 $t_0 = \sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} \cdot t_{\alpha/2,\nu}$, 여기서 $\sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} = \sqrt{MSE(1/n_1 + 1/n_2)}$

· $LSD = \sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} \cdot \mathrm{r}(2, \alpha, \nu)/\sqrt{2}$, where $i \neq j = 1, 2, \cdots, k$;

$t_{\nu,\alpha/2} = \mathrm{r}(2, \alpha, \nu)/\sqrt{2}$

where $\mathrm{r}(2, \alpha, \nu)$;유의수준 α에서 자유도가 ν이고
비교차수가 2인 경우의 표준화 범위값을 의미

· $|\overline{y_i} - \overline{y_j}| > LSD$ 이면 유의한 차이 존재

③ 특징; i) 실험전 미리 결정된 두 수준 비교에 매우 유용

ii) 각 수준의 반복 수가 동일한 경우에 유용

2) Tukey's Range Procedure (The HSD)

① 개념 : LSD와 동일하지만 LSD에서 이용하는 비교 차수 2 대신에 수준 수 k를 차수로 표준화하여 계산된 하나의 기각값을 이용하여 비교하는 방법

② 표현식

· $TK(=HSD) = \sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} \cdot \mathrm{r}(k, \alpha, \nu) / \sqrt{2}$

여기서 $i \neq j = 1, 2, \cdots, k$;

$t_{\nu, \alpha/2} = \mathrm{r}(k, \alpha, \nu) / \sqrt{2}$

$\mathrm{r}(k, \alpha, \nu)$;유의수준 α에서 자유도가 ν이고 수준 수가 k인 경우 표준화 분포를 따르는 표준화 범위값을 의미

· $|\overline{y_i} - \overline{y_j}| > HSD$ 이면 유의한 차이 존재

③ 특징

i) 수정된 LSD 방법으로 기본적인 과정은 LSD와 동일

ii) 표준화 차수를 비교 가능한 최대 차수인 수준 수로 이용하기 때문에 보다 큰 기각역을 제공하게 되어 LSD보다 더 보수적인 결과를 제공

iii) LSD와 마찬가지로 하나의 기각값을 이용

iv) 각 수준의 반복 수가 동일한 경우에 유용

3) Student-Newman-Keuls Multiple Range Test (The SNK)

① 개념 : 표준화 차수로 각 수준의 평균 순위차(p)를 이용하며, 동일한 유의수준의 가정에서 평균 순위차에 따라 기각값을 다르게 설정하여 수준간 평균 차이를 비교하는 방법

② 표현식

· $SNK(p) = \sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} \cdot \mathrm{r}(p, \alpha, \nu) / \sqrt{2}$

여기서 $p = 2,3,\cdots,k$; 평균 순위 차$(= i - j + 1)$, $i \neq j = 1,2,\cdots,k$;

$t_{\nu,\alpha/2} = \mathrm{r}(p,\alpha,\nu) / \sqrt{2}$

$\mathrm{r}(p,\alpha,\nu)$;유의수준 α에서 자유도가 ν이고 평균 순위 차가 p인 경우의 표준화범위값을 의미

· $|\overline{y_i} - \overline{y_j}| > SNK(p)$ 이면 유의한 차이 존재

③ 특징

ⅰ) 평균 순위 차마다 비교 그룹을 형성하여 각 비교 그룹마다 각기 다른 기각값을 설정

ⅱ) 일반적으로 큰 기각값을 제공하게 되어 상당히 보수적인 결과를 제공

4) Duncan's New Multiple Range Test (The NMRT)

① 개념 : 기본적인 개념과 방법은 SNK와 동일하지만 평균 순위 차에 따라 유의수준(α_p)을 다르게 설정하여 수준간 평균 차이를 비교하는 방법

② 표현식

·유의수준(α_p)의 정의; 유의수준 α; 비교 수준 p를 모두 기각시킬확률 이라고 하면

$$\alpha_p = 1 - (1-\alpha)^p$$
$= 1-$비교 수준 p를 모두 채택시킬신뢰수준의 평균
$=$ 비교 수준이 p인 경우 개별적으로 기각시킬 확률

· $D(p) = \sqrt{\widehat{Var}(\overline{y_i} - \overline{y_j})} \cdot \mathrm{r}^*(\mathrm{p}, \alpha_{\mathrm{p}}, \nu) / \sqrt{2}$

여기서 $p = 2,3,\cdots,k$; 평균 순위 차$(= i-j+1)$, $i \neq j = 1,2,\cdots,k$;
$t_{\nu,\alpha/2} = \mathrm{r}^*(p, \alpha_p, \nu) / \sqrt{2}$
$\mathrm{r}^*(p, \alpha_p, \nu)$;유의수준 α_p에서 자유도가 ν이고 평균 순위의 차가 p인 경우의 표준화범위값을 의미

· $|\overline{y_i} - \overline{y_j}| > D(p)$ 이면 유의한 차이 존재

③ 특징

i) SNK와 마찬가지로 수준 차이 비교시 평균 순위 차마다 각기 다른 기각값을 설정하여 비교

ii) 일반적으로 큰 기각값을 제공하여 비교적 보수적인 결과를 제공하므로 널리 이용

3. Inference about Structured Means

1) Scheffe's SCI

· 반복 수에 관계없이 사용 가능한 방법으로 SCI가 가장 짧은 구간을 제공

· $\sum_i c_i \overline{y_i} \pm \sqrt{(k-1)\,F_\alpha\,MSE \sum_i (c_i^{\,2}/r_i)}$

여기서 F_α : 자유도가 $(k-1), \nu$일 때의
상위 α% 에 해당하는 $F-$분포값

· $F_C = \sqrt{(k-1)\,F_\alpha\,MSE \sum_i (c_i^{\,2}/r_i)}$ 를 계산한 후, $|\sum_i c_i \overline{y_i}| > F_C$ 이면 유의한 차이 존재

2) Bonferroni's SCI

· 분석을 원하는 m 개의 대비를 가정할 때 유의수준을 $\alpha_m = \alpha / m$ 으로 사용하는 방법

· $\sum_i c_i \overline{y_i} \pm t_{\nu,\,\alpha/2m} \sqrt{MSE \sum_i (c_i^{\,2}/r_i)}$

· $BON = t_{\nu,\alpha/2m} \sqrt{MSE \sum_i (c_i^{\,2}/r_i)}$ 을 계산한 후, $|\sum_i c_i \overline{y_i}| > BON$ 이면 유의한 차이 존재